Multimedia Systems
Lecture – 22

By

Dr. Priyambada Subudhi

Assistant Professor

IIIT Sri City

Lossless Compression

- Lossless compression techniques achieve compression by removing the redundancy in the signal.
- This is normally achieved by assigning new codes to the symbols based on the frequency of occurrence of the symbols in the message.
- More frequent symbols are assigned shorter codes and vice versa.
- Sometimes the entire message to be coded is not readily available and frequency of symbols cannot be computed a priori.
- In such cases, a probabilistic model is assumed and the source symbols are assigned a model probability. The lossless compression algorithms make use of the probabilistic models to compute efficient codes for the symbols.
- lossless coding techniques based on a probabilistic model might not necessarily guarantee compression when the distribution of symbols in the message does not confirm to the probabilistic model used.

Run Length Encoding

- Instead of assuming a memoryless source, run-length coding (RLC) exploits memory present in the information source.
- It is one of the simplest forms of data compression.
- The basic idea is that if the information source we wish to compress has the property that symbols tend to form continuous groups, instead of coding each symbol in the group individually, we can code one such symbol and the length of the group.
- A sample string of symbols is as follows:

  BBBBEEEEEEEECCCCDAAAAA.
- It can be represented as

  4B8E4C1D5A.

- Run length encoding is used in a variety of tools involving text, audio, images, and video.
- Consider a bilevel image (one with only 1-bit black and white pixels) with monotone regions. This information source can be efficiently coded using run-length coding. In fact, since there are only two symbols, we do not even need to code any symbol at the start of each run. Instead, we can assume that the starting run is always of a particular color (either black or white) and simply code the length of each run.

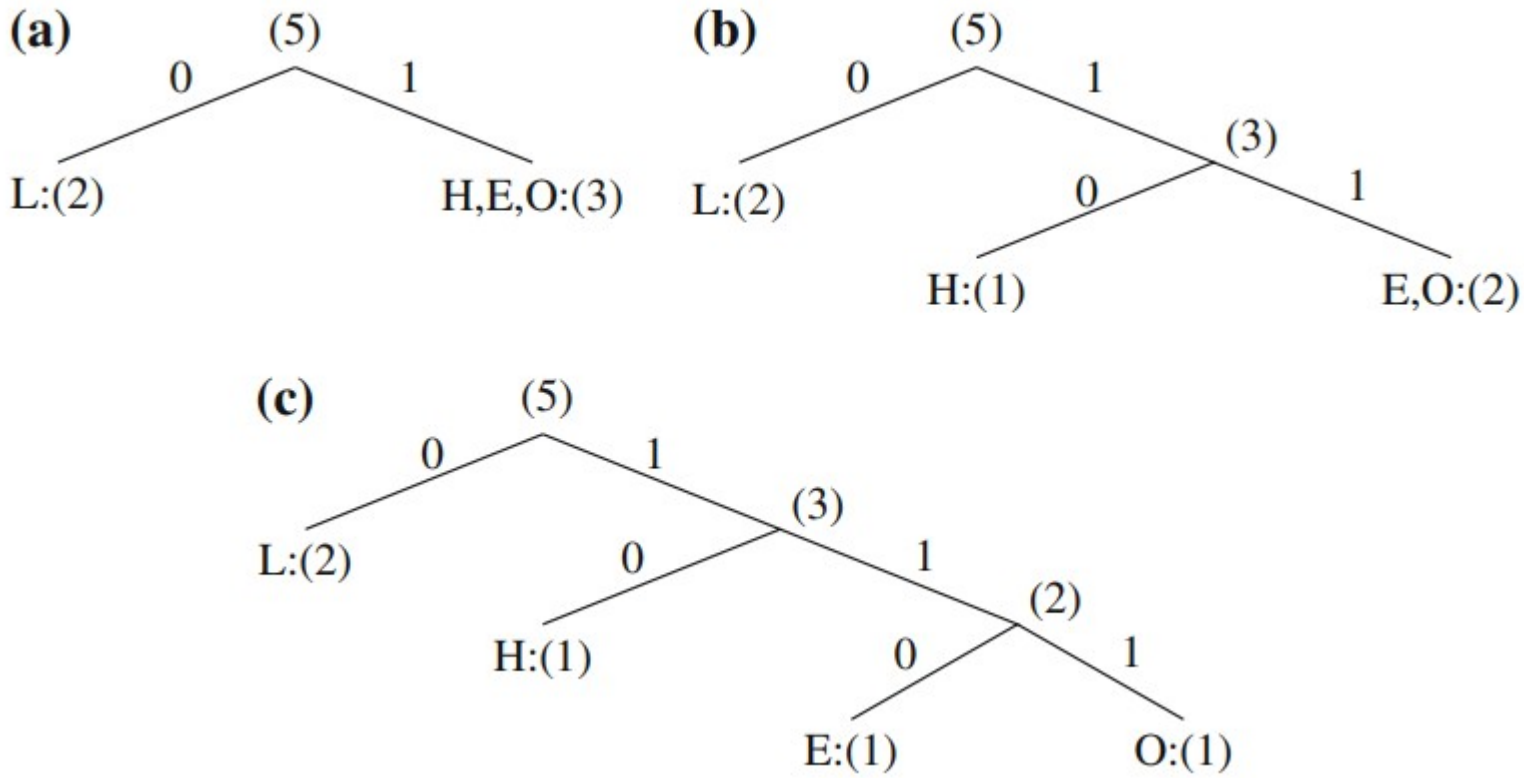
Variable-Length Coding

- Since the entropy indicates the information content in an information source S, it leads to a family of coding methods commonly known as entropy coding methods.
- Variable-length coding (VLC) is one of the best-known such methods.
- We will study the
  - Shannon–Fano algorithm
  - Huffman coding

Shannon–Fano Algorithm

- let us suppose the symbols to be coded are the characters in the word HELLO. The frequency count of the symbols is
- Symbol H E L O
- Count 1 1 2 1
- The encoding steps of the Shannon–Fano algorithm can be presented in the following top-down manner:
  - Sort the symbols according to the frequency count of their occurrences.
  - Recursively divide the symbols into two parts, each with approximately the same number of counts, until all parts contain only one symbol.

- Initially, the symbols are sorted as LHEO



One result of performing the Shannon–Fano algorithm on HELLO

Symbol	Count	$\log_2 \frac{1}{p_i}$	Code	Number of bits used
L	2	1.32	0	2
H	1	2.32	10	2
E	1	2.32	110	3
O	1	2.32	111	3
TOTAL number of bits:				10

- Entropy is given by

$$\eta = p_L \cdot \log_2 \frac{1}{p_L} + p_H \cdot \log_2 \frac{1}{p_H} + p_E \cdot \log_2 \frac{1}{p_E} + p_O \cdot \log_2 \frac{1}{p_O}$$
$$= 0.4 \times 1.32 + 0.2 \times 2.32 + 0.2 \times 2.32 + 0.2 \times 2.32 = 1.92$$